

Achieving the Capacity of the N -Relay Gaussian Diamond Network Within $\log N$ Bits

Bobbie Chern and Ayfer Özgür
Stanford University
Stanford, California 94305
{bgchern, aozgur}@stanford.edu

Abstract—We consider the N -relay Gaussian diamond network where a source node communicates to a destination node via N parallel relays. We show that several strategies can achieve the capacity of this network within $O(\log N)$ bits independent of the channel configurations and the operating SNR. The first of these strategies is partial decode-and-forward: the source node broadcasts independent messages to the relays at appropriately chosen rates, which in turn decode and forward these messages to the destination over a MAC channel. The same performance can be also achieved by compress-and-forward, quantize-map-and-forward or noisy network coding if relays quantize their observations at a decreasing resolution with N , instead of quantizing at the noise-level. The best capacity approximations currently available for this network are within $O(N)$ bits which follow from the corresponding capacity approximations for general Gaussian relay networks.

I. INTRODUCTION

Consider a Gaussian relay network where a source node communicates to a destination with the help of N relays. Characterizing the capacity of this network is a long-standing open problem in network information theory. Recently, progress has been made by showing that several strategies can achieve the capacity within a bounded gap: quantize-map-and-forward in [1], noisy network coding in [2], and compress-and-forward in [3]. The performance gap to capacity is independent of the channel parameters, the operating SNR, and the network topology. This makes these strategies universally good for relaying across different channel configurations, SNR regimes and network topologies. However, the gap increases linearly in the number of relay nodes N (or the total number of transmit and receive antennas when nodes are equipped with multiple antennas). This limits the applicability of these strategies to small networks with few relays.

In this paper, we aim to understand the reasons for this $O(N)$ performance gap to capacity and whether it can be improved. We consider the N -relay Gaussian diamond network, where the source node communicates to the destination in two stages. In the first stage, the source broadcasts its signal to the N relay nodes. In the second stage, the relays transmit their signals to the destination which observes a superposition of all the transmitted signals. See Fig. 1. Our goal is to understand

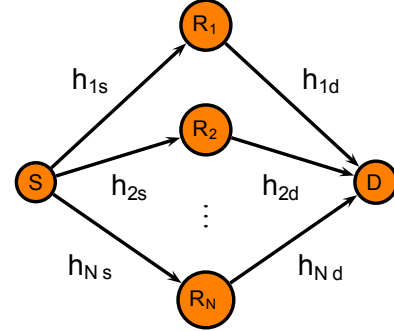


Fig. 1. The Gaussian N -relay diamond network.

if the strategies above can achieve the capacity of the N -relay diamond network within $O(\log N)$ bits instead of $O(N)$, independent of the channel parameters and the operating SNR, and if there are any other strategies that can yield the same performance.

We first show that a simple partial decode-and-forward strategy can achieve the N -relay diamond network capacity within $O(\log N)$ bits. Here, the source uses superposition coding to transmit independent messages to each of the relays at appropriately chosen rates; relays decode their intended messages from the source, re-encode and forward them to the destination over the multiple-access channel. We prove this result by showing that for all channel configurations and SNR, there exists a rate point in the intersection of the broadcast and multiple access capacity regions such that the sum rate is at most $O(\log N)$ bits away from the information theoretic cutset upper bound on the capacity of the network. We next show that the same performance can be achieved by quantize-map-and-forward, noisy network coding, and compress-and-forward. However, to achieve the capacity within $O(\log N)$ bits instead of $O(N)$ with these strategies, the relays have to quantize their observations at a resolution inversely proportional to N . In other words, the power of the quantization noise at each relay has to increase linearly in N . In [1], [2], [3], quantization is performed at the noise level. This shows that the rate penalty for describing the quantized observations can be significantly larger than the rate penalty associated with coarser quantization.

Capacity approximations for the N -relay diamond network have been also considered in [5] and [6]. [5] provides a hybrid

The work of Bobbie Chern was supported by the Department of Defense (DoD) through the National Defense Science & Engineering Graduate Fellowship (NDSEG) Program.

approximation for the capacity with both additive and multiplicative gaps by using only a subset of the available relays. [6] provides a constant gap approximation to the capacity, however only for the special case when all channel coefficients in the network are equal to each other. The approximation is based on an amplify-and-forward strategy at the relays. When channel coefficients are unequal, the performance of amplify-and-forward (also called analog network coding) can be arbitrarily away from capacity. For example, [5] shows that the best rate that can be achieved with amplify-and-forward in any N -relay diamond network is approximately equal to the rate achieved by using only the best relay, which can in turn be as small as half the capacity of the whole network. Therefore, amplify-and-forward can not provide constant gap approximations to capacity across different channel parameters and SNR's, such as the $O(\log N)$ approximation provided by the earlier strategies.

II. MODEL

We consider the Gaussian N -relay diamond network depicted in Fig. 1, where the source node S wants to communicate to the destination node D with the help of N relay nodes, denoted $\mathcal{N} = \{1, \dots, N\}$. All nodes are equipped with a single transmit and receive antenna. Let $X_s[t]$ and $X_i[t]$ denote the signals transmitted by the source node and the relay node $i \in \mathcal{N}$ respectively at time instant $t \in \mathbb{N}$. Similarly, $Y_d[t]$ and $Y_i[t]$ denote the signals received by the destination node and the relay node i respectively. The transmitted signal $X_i[t]$ by relay i is a causal function of its received signal $Y_i[t]$. We have

$$\begin{aligned} Y_i[t] &= h_{is}X_s[t] + Z_i[t], \\ Y_d[t] &= \sum_{i=1}^N h_{id}X_i[t] + Z[t], \end{aligned}$$

where h_{is} denotes the complex channel coefficient between the source and relay node i , and h_{id} denotes the complex channel coefficient between the relay node i and the destination node. We assume that the channel coefficients are known at all the nodes. $Z_i[t]$ and $Z[t]$ are independent and identically distributed circularly symmetric Gaussian random variables of variance σ^2 . All nodes are subject to an average power constraint P and we define $\text{SNR} := P/\sigma^2$. Note that the equal power constraint assumption is without loss of generality as the channel coefficients are arbitrary.

III. MAIN RESULT

The main conclusions of this paper are summarized in the following theorems.

Theorem 3.1. *Let \bar{C} be the information-theoretic cutset upper bound on the capacity of the N -relay diamond network. Then a partial decode-and-forward strategy at the relays achieves a rate*

$$R_{PDF} \geq \bar{C} - G_1, \quad (1)$$

where $G_1 = 2 \log N$.

Theorem 3.2. *Noisy network coding at the relays can achieve a rate*

$$R_{NNC} \geq \bar{C} - G_2, \quad (2)$$

where $G_2 = \log(N+1) + \log N + 1$. The same performance can be achieved by quantize-map-and-forward or compress and forward.

Remark The results can be extended to the case when nodes are equipped with multiple antennas. In this case, the gap is logarithmic in the total number of antennas at the relays.

IV. PARTIAL DECODE AND FORWARD

We consider a partial decode and forward strategy where the first stage of the communication is treated as a broadcast channel and the second stage is treated as a multiple access channel. The source uses superposition coding to send independent messages to the relays at rates R_i , $i \in \mathcal{N}$ that lie in the intersection of the broadcast and multiple access capacity regions. The relays decode these messages and re-encode and forward them to the destination over the MAC channel. The rate achieved by this strategy is given by

$$R_{PDF} = \sum_{i \in \mathcal{N}} R_i \quad \text{if} \quad \{R_1, \dots, R_N\} \in \mathcal{C}_{BC} \cap \mathcal{C}_{MAC}$$

where \mathcal{C}_{BC} and \mathcal{C}_{MAC} are the capacity regions of the BC channel at the first stage and the MAC channel at the second stage respectively. Therefore, in order to bound the gap between the rate achieved by this strategy and the information theoretic cutset upper bound on the capacity of the network, we need to show that there exists a rate point $\{R_1, \dots, R_N\} \in \mathcal{C}_{BC} \cap \mathcal{C}_{MAC}$ such that the difference between $\sum_{i \in \mathcal{N}} R_i$ and \bar{C} is bounded. We use ideas on polymatroid intersection inspired by [3], which uses submodular flows to characterize the binning rates in a compress-and-forward relaying strategy.

The region \mathcal{C}_{MAC} is known to have a polymatroid structure [8]. The region \mathcal{C}_{BC} however is not polymatroidal. Below, we define a polymatroid, and use the duality between the BC and MAC capacity regions [7] to find a polymatroidal lower bound on the BC capacity region. We then use Edmond's polymatroid intersection ([9], Corollary 46.1b) to find an intersection point in the two polymatroid regions with largest sum capacity.

Definition 4.1. Let $f : 2^{\mathcal{N}} \rightarrow \mathbb{R}^+$ be a set function. The polyhedron

$$P(f) := \{(x_1, \dots, x_N) : \sum_{i \in S} x_i \leq f(S), \forall S \subseteq \mathcal{N}, x_i \geq 0, \forall i\}$$

is a polymatroid if the set function f satisfies

- 1) $f(\emptyset) = 0$ (normalized).
- 2) $f(S) \leq f(T)$ if $S \subseteq T$ (non-decreasing).
- 3) $f(S) + f(T) \geq f(S \cup T) + f(S \cap T)$ (submodular).

The MAC capacity region \mathcal{C}_{MAC} is given by

$$P(f) = \{(R_1, \dots, R_N) : \sum_{i \in S} R_i \leq f(S), S \subseteq \mathcal{N}, R_i \geq 0, \forall i\}$$

where

$$f(S) = \log \left(1 + \sum_{i \in S} |h_{id}|^2 \text{SNR} \right).$$

Since f satisfies the conditions in Definition 4.1, $P(f)$ is a polymatroid. By the duality established in [7], the BC capacity region is given by

$$C_{BC} = \bigcup_{(P_1, \dots, P_N) : \sum P_i = P} \mathcal{C}_{MAC}(P_1, \dots, P_N)$$

where $\mathcal{C}_{MAC}(P_1, \dots, P_N)$ is the capacity region of the MAC channel from the relays to the source node with relay i transmitting at power P_i . In particular, $\mathcal{C}_{MAC}(P/N, \dots, P/N) \subseteq C_{BC}$, i.e.,

$$C_{BC} \supseteq P(g)$$

where

$$g(S) = \log \left(1 + \sum_{i \in S} |h_{is}|^2 \frac{\text{SNR}}{N} \right).$$

Clearly, $P(g)$ is also a polymatroid. It then follows from Edmond's polymatroid intersection ([9], Corollary 46.1b) that

$$\begin{aligned} \max \left\{ \sum_i R_i : (R_1, \dots, R_N) \in P(f) \cap P(g) \right\} \\ = \min_{\Lambda \subseteq N} f(\bar{\Lambda}) + g(\Lambda). \end{aligned}$$

Therefore, partial decode and forward can achieve a rate

$$\begin{aligned} R_{PDF} &= \min_{\Lambda \subseteq N} g(\bar{\Lambda}) + f(\Lambda) \\ &= \min_{\Lambda \subseteq N} \left(\log \left(1 + \sum_{i \in \bar{\Lambda}} |h_{is}|^2 \text{SNR}/N \right) \right. \\ &\quad \left. + \log \left(1 + \sum_{i \in \Lambda} |h_{id}|^2 \text{SNR} \right) \right) \end{aligned} \quad (3)$$

This result can be also obtained as a special case of Theorem 5 in [4].

We now bound the gap between this achievable rate and the information theoretic cutset upper bound on the capacity of the network. The cutset upper is given by [10]

$$\bar{C} = \sup_{X, X_1, \dots, X_N} \min_{\Lambda \subseteq N} I(X, X_\Lambda; Y, Y_{\bar{\Lambda}} | X_{\bar{\Lambda}}),$$

where $X_\Lambda = \{X_i, i \in \Lambda\}$ and $Y_{\bar{\Lambda}}, X_{\bar{\Lambda}}$ are defined analogously. Replacing the order of maximization and minimization, the cutset upper bound can be upper bounded as

$$\begin{aligned} \bar{C} &\leq \min_{\Lambda \subseteq N} \left(\log \left(1 + \text{SNR} \sum_{i \in \bar{\Lambda}} |h_{is}|^2 \right) \right. \\ &\quad \left. + \log \left(1 + \text{SNR} \left(\sum_{i \in \Lambda} |h_{id}|^2 \right) \right) \right) \end{aligned} \quad (4)$$

It can be easily verified that

$$R_{PDF} \geq \bar{C} - 2 \log N.$$

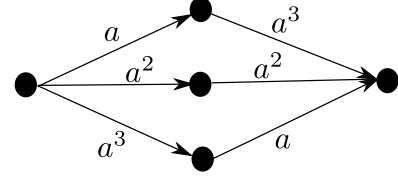


Fig. 2. A 3-relay diamond network.

A. Discussion

Consider the 3-relay diamond network in Fig. 2 where the labels indicate the SNR's of the corresponding links with the transmit and noise powers normalized to 1. Considering the deterministic model of [1] for this network suggests that each relay should carry information at rate approximately $\log a$ when a is large. If we use partial decode-and-forward at the relays, one natural choice for the powers of the superposed codebooks at the source intended for different relays can be $P_1 = 1 - 1/a - 1/a^2$, $P_2 = 1/a$, and $P_3 = 1/a^2$. At large a , this corresponds to communication rates

$$R_1 = \log \left(1 + \frac{aP_1}{1 + a(P_2 + P_3)} \right) \approx \log a - 1$$

$$R_2 = \log \left(1 + \frac{a^2 P_2}{1 + a^2 P_3} \right) \approx \log a - 1$$

$$R_3 = \log(1 + a^3 P_3) \approx \log a$$

to the three relays. There is a 1 bit/s/Hz rate loss at each relay (except for the strongest one) since the codebooks intended for the stronger relays constitute additional noise at the weaker relays. In the corresponding extension of this configuration to N -relays, this would result in $O(N)$ rate loss between the sum broadcast rate to the relays and the capacity of the single-input multiple output (SIMO) channel at the first stage, i.e the cut-set upper bound. The above argument suggests that there is a better way to choose the broadcasting rates to the relays. Instead, we can choose $P_1 = 1 - \sum_{i=2}^N P_i$ and $P_i = \frac{N-i+1}{a^{i-1}}$ for $i = 1, \dots, N$ and we obtain the rates

$$R_1 \approx \log a - \log N, \quad R_2 \approx \log a, \quad \dots \quad R_N \approx \log a.$$

which also lie in the broadcast capacity region of the first stage. In this case, the sumrate is only $O(\log N)$ bits/s/Hz away from the SIMO capacity.

V. QUANTIZE-MAP-AND-FORWARD, NOISY NETWORK CODING, COMPRESS-AND-FORWARD

In this section, we investigate the performance of quantize-map-and-forward, noisy network coding and compress-and-forward relaying strategies. The three strategies are strongly related as they involve the same basic operation at the relays: quantizing the received signal and independently mapping it to a transmit codeword either with or without binning. We take the noisy network coding result in [2] as a reference. Noisy network coding reduces to quantize-map-and-forward in the context of the diamond network. It has been shown in [4], [11] that the same performance can be achieved by

compress-and-forward, where the quantized signals are binned before transmission at the relays at appropriately chosen rates, and they are decoded successively before decoding the actual source message.

The performance achieved by noisy network coding is given in [2, Theorem 1] as

$$R_{NNC} = \min_{\Lambda \subseteq \mathcal{N}} I(X_s, X_\Lambda; Y_d, \hat{Y}_{\bar{\Lambda}} | X_{\bar{\Lambda}}) - I(Y_\Lambda; \hat{Y}_\Lambda | X, X_{\mathcal{N}}, \hat{Y}_{\bar{\Lambda}}, Y_d). \quad (5)$$

for some joint probability distribution $\prod_{i \in \mathcal{N}} p(x_i) p(\hat{y}_i | y_i, x_i)$. Choosing X_i to be i.i.d. circularly symmetric Gaussian of variance P and

$$\hat{Y}_i = Y_i + \hat{Z}_i, \quad i \in \mathcal{N},$$

where $\hat{Z}_i, i \in \mathcal{N}$ are i.i.d. circularly symmetric and complex Gaussian random variables of variance $N\sigma^2$, the first mutual information becomes

$$\begin{aligned} I(X_s, X_\Lambda; Y_d, \hat{Y}_{\bar{\Lambda}} | X_{\bar{\Lambda}}) &= \log \left(1 + \sum_{i \in \bar{\Lambda}} |h_{is}|^2 \text{SNR} / (N+1) \right) \\ &\quad + \log \left(1 + \sum_{i \in \Lambda} |h_{id}|^2 \text{SNR} \right), \end{aligned}$$

since the quantized observations are corrupted by quantization and thermal noise with total variance $(N+1)\sigma^2$. Note that the expression is similar to the rate achieved in (3). The second term is given by

$$I(Y_\Lambda; \hat{Y}_\Lambda | X, X_{\mathcal{N}}, \hat{Y}_{\bar{\Lambda}}, Y_d) = |\Lambda| \log \left(1 + \frac{1}{N} \right) \leq \frac{|\Lambda|}{N},$$

Therefore the total gap of (5) to the cutset-upper bound in (4) is bounded by $\log(N+1) + \log N + 1$.

REFERENCES

- [1] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, *Wireless Network Information Flow: A Deterministic Approach*, IEEE Trans. Info. Theory, vol. 57, no. 4, pp. 1872-1905, 2011.
- [2] S. H. Lim, Y.-H. Kim, A. El Gamal, S.-Y. Chung, *Noisy Network Coding*, IEEE Trans. Info. Theory, vol. 57, no. 5, pp. 3132-3152, May 2011.
- [3] A. Raja and P. Viswanath, *Compress-and-Forward Scheme for a Relay Network: Approximate Optimality and Connection to Algebraic Flows*, IEEE Int. Symposium on Information Theory (ISIT) St Petersburg, 2011; e-print <http://arxiv.org/abs/1012.0416>.
- [4] S. Kannan and P. Viswanath, *Capacity of Multiple Unicast in Wireless Networks: A Polymatroidal Approach*, IEEE Int. Symposium on Information Theory (ISIT), St Petersburg, 2011; e-print <http://arxiv.org/abs/1111.4768>.
- [5] C. Nazeroglu, A. Özgür, and C. Fragouli, *Wireless Network Simplification: the Gaussian N-Relay Diamond Network* IEEE Int. Symposium on Information Theory (ISIT), St Petersburg, 2011; e-print <http://arxiv.org/abs/1103.2046>.
- [6] U. Niesen, S. Diggavi, *The Approximate Capacity of the Gaussian N-Relay Diamond Network*, IEEE Int. Symposium on Information Theory (ISIT), St Petersburg, 2011.
- [7] S. Vishwanath, N. Jindal, and A. J. Goldsmith, *Duality, achievable rates and sum-rate capacity of Gaussian MIMO broadcast channel*, IEEE Trans. Info. Theory, vol. 49, pp. 2658-2668, 2003.
- [8] D. Tse and S. Hanly, *Multiaccess Fading Channels Part I: Polymatroid Structure, Optimal Resource Allocation and Throughput Capacities*, vol. 44, no. 7, pp. 2786-2816, November 1998.
- [9] A. Schrijver, *Combinatorial Optimization*, Springer, Berlin, 2003.
- [10] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd Edition. John Wiley and Sons, New York, 2006.
- [11] X. Wu and L.-L. Xie, *On the Optimal Compressions in the Compress-and-Forward Relay Schemes*, e-print <http://arxiv.org/abs/1009.5959>.